

THE DENOMINATORS OF HARMONIC NUMBERS

PETER SHIU

ABSTRACT. The denominators d_n of the harmonic number $1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ do not increase monotonically with n . It is conjectured that $d_n = D_n = \text{LCM}(1, 2, \dots, n)$ infinitely often. For an odd prime p , the set $\{n : pd_n | D_n\}$ has a harmonic density and, for $2 < p_1 < p_2 < \cdots < p_k$, there exists n such that $p_1 p_2 \cdots p_k d_n | D_n$.

1. INTRODUCTION

Although much is known concerning the asymptotic behaviour of the harmonic number

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n},$$

there is a dearth of results on the number itself as a fraction. Let c_n, d_n and D_n be defined by

$$H_n = \frac{c_n}{d_n}, \quad \text{GCD}(c_n, d_n) = 1; \quad D_n = \text{LCM}(1, 2, \dots, n);$$

the following is a small table for their values:

n	1	2	3	4	5	6	7	8	9	10
c_n	1	3	11	25	137	49	363	761	7129	7381
d_n	1	2	6	12	60	20	140	280	2520	2520
D_n	1	2	6	12	60	60	420	840	2520	2520

Table 1

Wolstenholme's theorem [4, Theorem 115] states that $p^2 | c_{p-1}$ for primes $p > 3$. More recently, A. Eswarathasan and E. Levine [3] conjectured that each odd prime p divides only finitely many c_n , and D. W. Boyd [1] has given computation results and heuristic arguments based on p -adic analysis to support the conjecture. If the conjecture is true, there exists $N = N(p)$ such that $p \nmid c_m$ for all $m \geq N$; now, for $n \geq pN$, we may set $n = mp + r$, $0 \leq r < p$, so that there are precisely m terms in the sum for H_n in which the denominator is a multiple of p , with their sum being H_m/p , and the denominator of the reduced fraction for $H_n - H_m/p$ is free of the prime p . It then follows from $p \nmid c_m$ that $p | d_n$ for $n \geq pN$.

The sequences (c_n) and (d_n) are not monotonic, and we prove the following:

Date: 5 July 2016.

1991 *Mathematics Subject Classification.* Primary (2010): 11D68; Secondary (2010): 11Y70.

Key words and phrases. Harmonic density, Kronecker's theorem, Wieferich primes.

Theorem 1. *Each of the following holds for infinitely many n :*

(i) $d_n > d_{n-1}$, (ii) $d_n = d_{n-1}$, (iii) $d_n < d_{n-1}$; (iv) $c_n > c_{n-1}$, (v) $c_n < c_{n-1}$; and $c_n \neq c_{n-1}$ for all $n > 1$.

It is easy to show that the exponent of 2 in the prime factorisation of d_n is the same as that in D_n . Write

$$D_n = d_n q_n,$$

so that c_n and q_n are odd numbers; and, for odd primes p , define the sets

$$\mathcal{E}_p = \{n : 1 < n < p, p|c_n\}, \quad \mathcal{Q}_p = \{n : p|q_n\}.$$

Summing the identity

$$(1) \quad \frac{1}{j} + \frac{1}{p-j} = \frac{p}{j(p-j)}, \quad 1 \leq j \leq \frac{p-1}{2},$$

over j , the left-hand side delivers H_{p-1} , so that $p|c_{p-1}$, which is enough for our purpose without appealing to Wolstenholme's theorem. Anyway, $p-1 \in \mathcal{E}_p$, and we have the following theorems concerning \mathcal{Q}_p :

Theorem 2. *Let $m \in \mathcal{E}_p$. Then $n \in \mathcal{Q}_p$ for each n satisfying*

$$(2) \quad mp^a \leq n < (m+1)p^a, \quad a = 1, 2, \dots$$

Conversely, if $n \in \mathcal{Q}_p$ then (2) holds for some $m \in \mathcal{E}_p$ and $a \geq 1$.

Theorem 3. *The set \mathcal{Q}_p has the harmonic density*

$$\delta(\mathcal{Q}_p) = \frac{1}{\log p} \sum_{m \in \mathcal{E}_p} \log \left(1 + \frac{1}{m}\right).$$

Theorem 4. *Let $2 < p_1 < p_2 < \dots < p_k$ be prime numbers. Then there exists q_n such that $p_1 p_2 \dots p_k | q_n$.*

Since $q_n = 1$ means that $n \notin \mathcal{Q}_p$ for all $p \geq 3$, we also define

$$\tilde{\mathcal{Q}} = \text{Comp} \bigcup_{p \geq 3} \mathcal{Q}_p, \quad \tilde{\mathcal{Q}}(x) = \sum_{n \leq x, n \in \tilde{\mathcal{Q}}} 1,$$

so that $d_n = D_n$ if and only if $n \in \tilde{\mathcal{Q}}$. It seems likely that there are infinitely many prime powers $p^a \in \tilde{\mathcal{Q}}$, and we propose the following:

Conjecture. *There are positive constants K_1, K_2 such that*

$$\frac{K_1 x}{\log x} < \tilde{\mathcal{Q}}(x) < \frac{K_2 x}{\log x}, \quad x > 1.$$

2. PROOF OF THEOREM 1

(i) The last term in the sum for $H_p = c_p/d_p$, namely $1/p$, is the only term in which the denominator is a multiple of p . It follows that $p|d_p$, so that $d_p \geq p$ is unbounded, and hence $d_n > d_{n-1}$ for infinitely many n .

(ii) We prove that $d_n = d_{n-1}$ when $n = 2p > 6$. Again, $p|d_{n-1}$ because $1/p$ is the only term in the sum for H_{n-1} in which the denominator is a multiple of p . Next, the only such terms in the sum for H_n are $1/p$ and $1/2p$, with their sum being $3/p$;

since $p > 3$, we deduce that $p|d_n$. Indeed, with d_{n-1} and d_n being even, we further deduce that $n = 2p$ divides both of them. Finally, it follows from

$$\frac{c_n}{d_n} = \frac{c_{n-1} + d_{n-1}/n}{d_{n-1}} \quad \text{and} \quad \frac{c_{n-1}}{d_{n-1}} = \frac{c_n - d_n/n}{d_n}$$

that $d_n|d_{n-1}$ and $d_{n-1}|d_n$, so that $d_n = d_{n-1}$.

(iii) We prove that $d_n < d_{n-1}$ when $n = p(p-1)$. There are precisely $p-2$ terms in the sum for H_{n-1} in which the denominators are multiples of p ; in fact

$$H_{n-1} = \frac{H_{p-2}}{p} + S_0 + S_1 + \cdots + S_{p-2},$$

where

$$S_i = \frac{1}{ip+1} + \frac{1}{ip+2} + \cdots + \frac{1}{ip+p-1}, \quad i = 0, 1, 2, \dots, p-2,$$

and an obvious generalisation of (1) shows that the numerator of the reduced fraction for each S_i is a multiple of p . From

$$H_{n-1} = \frac{c_{n-1}}{d_{n-1}}, \quad \frac{H_{p-2}}{p} = \frac{1}{p} \left(H_{p-1} - \frac{1}{p-1} \right) = \frac{c_{p-1}}{pd_{p-1}} - \frac{1}{n},$$

we now find that

$$\frac{c_n}{d_n} = \frac{c_{n-1}}{d_{n-1}} + \frac{1}{n} = \frac{c_{p-1}}{pd_{p-1}} + S_0 + S_1 + \cdots + S_{p-2}.$$

Since $p|c_{p-1}$, the denominator of the reduced fraction on the right-hand side of the equation is free of the prime p , so that $p \nmid d_n$, and also $p|d_{n-1}$, because $p|n$. Finally, from $d_n|nd_{n-1}$, we now have $p^2 d_n|nd_{n-1}$, so that $d_n \leq nd_{n-1}/p^2 < d_{n-1}$.

We do not require $p|c_n$ when $p > 3$, which follows from Wolstenholme's theorem.

(iv) and (v) If $d_n \geq d_{n-1}$ then $c_n = H_n d_n > H_{n-1} d_{n-1} = c_{n-1}$, and if $d_n < d_{n-1}$ then

$$c_n = d_n \left(\frac{c_{n-1}}{d_{n-1}} + \frac{1}{n} \right) < \frac{d_n c_{n-1}}{d_{n-1}} < c_{n-1}.$$

The required results, including $c_n \neq c_{n-1}$, follow from (i), (ii) and (iii). \square

3. PROOF OF THEOREM 2

Take any $m \in \mathcal{E}_p$, and let n satisfy (2). Then $n > p^a$ so that $p^a|D_n$, and we have

$$H_n = \frac{H_m}{p^a} + \sum_{\substack{k \leq n \\ p^a \nmid k}} \frac{1}{k} = \frac{c_m/p}{p^{a-1}d_m} + \sum_{\substack{k \leq n \\ p^a \nmid k}} \frac{1}{k}.$$

From $p|c_m$, and hence $p \nmid d_m$, it follows that $p^a \nmid d_n$, so that $p|q_n$, that is $n \in \mathcal{Q}_p$.

Conversely, let $n \in \mathcal{Q}_p$, and let a and m be defined by $p^a \leq n < p^{a+1}$ and $m = \lfloor n/p^a \rfloor$, so that $1 \leq m < p$, and n satisfies (2). If $m \notin \mathcal{E}_p$ then $p^a|d_n$, and since $p^{a+1} \nmid D_n$, it follows that $p \nmid q_n$, contradicting $n \in \mathcal{Q}_p$. \square

4. PROOF OF THEOREM 3

The length of the interval (2) is p^a so that, for each $m \in \mathcal{E}_p$, the number of $n \in \mathcal{Q}_p$ with $n \leq x = p^b$ is $p + p^2 + \cdots + p^{b-1} = (p^b - p)/(p - 1)$. The intervals corresponding to different $m \in \mathcal{E}_p$ are disjoint, so that

$$(3) \quad \mathcal{Q}_p(x) = \frac{|\mathcal{E}_p|(p^b - p)}{p - 1} \sim \frac{|\mathcal{E}_p|x}{p} \quad \text{as } x = p^b \rightarrow \infty.$$

However, since members of \mathcal{Q}_p lie in long consecutive runs of integers, the set does not possess an asymptotic density. On the other hand its harmonic density exists. As $x \rightarrow \infty$, we have

$$\sum_{x \leq n < y} \frac{1}{n} = \log \frac{y}{x} + O\left(\frac{1}{x}\right),$$

so that

$$\sum_{mp^a \leq n < (m+1)p^a} \frac{1}{n} = \log \frac{m+1}{m} + O\left(\frac{1}{mp^a}\right),$$

and hence

$$\sum_{m \in \mathcal{E}_p} \sum_{1 \leq a < b} \sum_{mp^a \leq n < (m+1)p^a} \frac{1}{n} = b \sum_{m \in \mathcal{E}_p} \log \left(1 + \frac{1}{m}\right) + O(\log p).$$

For $x > 1$, we choose b so that $p^b \leq x < p^{b+1}$, and

$$\sum_{\substack{x < n < p^{b+1} \\ n \in \mathcal{Q}_p}} \frac{1}{n} \leq \log \frac{p^{b+1}}{x} + O\left(\frac{1}{x}\right) \leq \log p + O\left(\frac{1}{x}\right).$$

It then follows that the harmonic density for \mathcal{Q}_p is given by

$$\delta(\mathcal{Q}_p) = \lim_{x \rightarrow \infty} \frac{1}{\log x} \sum_{\substack{n \leq x \\ n \in \mathcal{Q}_p}} \frac{1}{n} = \lim_{b \rightarrow \infty} \frac{1}{b \log p} \sum_{\substack{n < p^{b+1} \\ n \in \mathcal{Q}_p}} \frac{1}{n} = \frac{1}{\log p} \sum_{m \in \mathcal{E}_p} \log \left(1 + \frac{1}{m}\right). \quad \square$$

5. PROOF OF THEOREM 4

Take $m_i = p_i - 1 \in \mathcal{E}_{p_i}$, $i = 1, 2, \dots, k$ in Theorem 2. If n belongs to the intersection of the intervals (2) corresponding to m_i , then q_n is divisible by each p_i . The idea then is to find a suitable set of integers a_i so that the intervals can be aligned to deliver a non-empty intersection, or better still, to form a nested sequence of intervals. That such an alignment is possible follows from an application of Kronecker's theorem [4, Theorem 443] on simultaneous approximations.

Lemma 1. *Let $0 < \delta < 1$, and $2 < p_1 < p_2 < \cdots < p_k$ be prime numbers. Then there are integers $a_1 > a_2 > \cdots > a_k > 1$ such that*

$$(1 - \delta)p_1^{a_1} < p_{i+1}^{a_{i+1}} < p_i^{a_i} \leq p_1^{a_1}, \quad i = 1, 2, \dots, k-1.$$

Proof. Let

$$\theta_i = \frac{\log p_1}{\log p_i}, \quad i = 1, 2, \dots, k,$$

so that $1 = \theta_1, \theta_2, \theta_3, \dots, \theta_k$ are linearly independent. By Kronecker's theorem, corresponding to any $\epsilon > 0$, there are positive integers a_1, a_2, \dots, a_k such that

$$\frac{(i-1)\log(1+\epsilon)}{k\log p_i} \leq a_1\theta_i - a_i < \frac{i\log(1+\epsilon)}{k\log p_i},$$

so that

$$\frac{(i-1)\log(1+\epsilon)}{k} \leq a_1\log p_1 - a_i\log p_i < \frac{i\log(1+\epsilon)}{k},$$

and hence

$$a_1\log p_1 - \frac{i\log(1+\epsilon)}{k} < a_i\log p_i \leq a_1\log p_1 - \frac{(i-1)\log(1+\epsilon)}{k}.$$

It follows that the sequence $(a_i\log p_i)$ is decreasing for $i = 1, 2, \dots, k$, and bounded by $a_1\log p_1 - \log(1+\epsilon)$ and $a_1\log p_1$, that is

$$\frac{p_1^{a_1}}{1+\epsilon} < p_{i+1}^{a_{i+1}} < p_i^{a_i} \leq p_1^{a_1}, \quad i = 1, 2, \dots, k-1.$$

That $a_{i+1} < a_i$ is a consequence of $p_{i+1} > p_i$. The required result follows by setting $\epsilon = \delta/(1-\delta)$, so that $(1+\epsilon)(1-\delta) = 1$. \square

Proof of Theorem 4. Replace a_i by $a_i + 1$ in the lemma so that

$$(1-\delta)p_1^{a_1+1} < p_{i+1}^{a_{i+1}+1} < p_i^{a_i+1} \leq p_1^{a_1+1}, \quad i = 1, 2, \dots, k-1,$$

and hence

$$1-\delta = \frac{(1-\delta)p_1^{a_1+1}}{p_1^{a_1+1}} < \frac{p_{i+1}^{a_{i+1}+1}}{p_i^{a_i+1}} < 1.$$

Choosing δ to satisfy

$$1-\delta > \min_{1 \leq i < k} \frac{p_{i+1}(p_i-1)}{p_i(p_{i+1}-1)},$$

we find that

$$\frac{(p_i-1)p_i^{a_i}}{(p_{i+1}-1)p_{i+1}^{a_{i+1}}} = \frac{p_{i+1}(p_i-1)p_i^{a_i+1}}{p_i(p_{i+1}-1)p_{i+1}^{a_{i+1}+1}} < 1, \quad i = 1, 2, \dots, k-1,$$

so that

$$(p_i-1)p_i^{a_i} < (p_{i+1}-1)p_{i+1}^{a_{i+1}} \leq (p_k-1)p_k^{a_k} = (1-1/p_p)p_k^{a_k+1} < p_k^{a_k+1}.$$

With $m_i = p_i - 1 \in \mathcal{E}_{p_i}$, the k intervals

$$\mathcal{I}(i) = \{n : m_i p_i^{a_i} \leq n < (m_i + 1)p_i^{a_i}\}, \quad i = 1, 2, \dots, k,$$

form a nested sequence, with $\mathcal{I}(i+1) \subset \mathcal{I}(i)$. By (2) in Theorem 2, $p_i | q_n$ for every $n \in \mathcal{I}(i)$, so that $p_1 p_2 \cdots p_k | q_n$ for every $n \in \mathcal{I}(k)$. The theorem is proved. \square

6. FLUCTUATIONS IN d_n

The sequence (D_n) is increasing, and $D_{n-1} < D_n$ if and only if $n = p^a$, a prime power—it follows that there are arbitrarily long runs of D_n taking the same value. Indeed the inequalities $D_{n-1} < D_n < D_{n+1}$ imply that both n and $n+1$ are prime powers, so that $\{n, n+1\} = \{2^a, p^b\}$; for example, $n = 16$ or $n = 31$.

The situation for (d_n) is not so simple, and there are interesting problems related to Theorem 1. Thus we may ask if there are arbitrarily long runs of d_n which are strictly increasing, or strictly decreasing. For example, we find that $d_n < d_{n+1} < d_{n+2} < d_{n+3}$ when $n = 1, 2, 6, 22, 70, 820, 856, 1288$. On the other hand

we have found only one solution to $d_n > d_{n+1} > d_{n+2}$, namely $n = 19$; in fact $d_{19} = 2^4 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19$, and $d_{20} = d_{19}/5$, $d_{21} = d_{20}/3$.

A long interval for n in which D_n is stationary may include a long subinterval in which d_n is stationary. For example, D_n takes the same value in $1331 \leq n < 1361$; note that $1331 = 11^3$, 1361 is a prime, and that $1332 = (p-1)p$, with $p = 37$. The argument in §1 shows that, for $n = (p-1)p + r$, $0 \leq r < p$, the denominator of the reduced fraction for $H_n - H_{p-1}/p$ is free of p , and, since $p|c_{p-1}$, we deduce that $p \nmid d_n$ for $(p-1)p \leq n < p^2 = 1369$; in fact, we find that

$$D_n = D_{1331} = d_{1331} = 37d_n \quad \text{for } 1332 \leq n < 1361.$$

7. LISTING AND COUNTING $n \leq x$ WITH $d_n = D_n$

By the prime number theorem, $D_n = \exp((1 + o(1))n)$ as $n \rightarrow \infty$, so that, if n is not small, it is not feasible to compute H_n in order to check whether $d_n = D_n$. However, we can apply Theorem 2 to compute q_n and hence the prime decomposition of $d_n = D_n/q_n$. In particular, we can check whether $n \in \tilde{Q}$, which amounts to showing that n does not lie in any of the intervals (2) associated with \mathcal{Q}_p . If we wish to list $n \in \tilde{Q}$ in an interval $1 \leq n \leq x$, we can apply a process similar to the sieve of Eratosthenes for the listing of prime numbers. By this we mean that, for each $m \in \mathcal{E}_p$, each $p \leq x/m$ and each $a \leq \log x / \log p$, we delete, or sift-out, the integers n in the interval (2) from the interval $1 \leq n \leq x$; the un-sifted numbers n then satisfy $d_n = D_n$. Such $n \leq 10000$ are given in §8.

It appears that the set \tilde{Q} can be dealt with using methods applied to the set of primes, and our Conjecture is based on (3). For example, instead of listing the members of \tilde{Q} , we may only wish to evaluate, or to estimate, its counting function $\tilde{Q}(x)$ by applying the inclusion-exclusion principle, in the same way that one does for the prime counting function $\pi(x)$. The number of $n \leq x$ which are divisible by a has the simple formula $[x/a]$, but it is not so easy to derive a formula for the counting function associated with the intersection of various sets \mathcal{Q}_p . Indeed we have yet to discover why there are arbitrarily large n with $d_n = D_n$; in particular, we have not been able to emulate Euclid's elegant proof that there are infinitely many primes.

8. COMPUTATION RESULTS

There are 2641 numbers $n \leq 10000$ such that $d_n = D_n$; they lie in 26 runs of consecutive numbers—in the display below, a_b denotes the interval $a \leq n < a + b$.

1₅, 9₉, 27₆, 49₅, 88₁₂, 125₃₁, 243₂₉, 289₅, 361₂, 484₂, 841₆, 968₉₉, 1164₂,
1331₁, 1369₈₉, 2401₉₉, 3125₂₅₄, 3488₁₇₂, 3721₂₇₂, 6889₈₇, 7085₄₇₇, 7761₇₁,
7921₃₂₄, 8342₅₁, 8502₃₇₅, 9156₁₅₅.

Thus 1₅ being followed by 9₉ means that $1, 2, 3, 4, 5, 9, 10, \dots, 17 \in \tilde{Q}$ and $6, 7, 8 \notin \tilde{Q}$. For $18 \leq n < 27$, there are precisely two terms in the sum for H_n in which the denominators are divisible by 3^2 , namely $1/9$ and $1/18$, with their sum being $1/6$, so that $3^2 \nmid d_n$, whereas $3^2 | D_n$. The argument amounts to taking $p = 3$, $m = a = 2$ in (2) to show that such $n \in \mathcal{Q}_3$. We remark that 15 of the leading terms a for a run are prime powers.

There is symmetry for \mathcal{E}_p in that, for $m, m' > 0$ with $m + m' + 1 = p$, we have $m \in \mathcal{E}_p$ if and only if $m' \in \mathcal{E}_p$; the proof makes use of the identity (1). Since $0 \notin \mathcal{E}_p$ and $p - 1 \in \mathcal{E}_p$, it follows that $|\mathcal{E}_p|$ is odd, unless $(p - 1)/2 \in \mathcal{E}_p$. A result of Eisenstein [2] states that

$$-\frac{H_r}{2} \equiv \frac{2^{p-1} - 1}{p} \pmod{p}, \quad r = \frac{p-1}{2},$$

so that $(p - 1)/2 \in \mathcal{E}_p$ if and only if $2^p \equiv 2 \pmod{p^2}$. Such primes p are called Wieferich primes, with only two known ones: $p = 1093, 3511$, and we find that

$$\mathcal{E}_{1093} = \{273, 546, 819, 1092\}, \quad \mathcal{E}_{3511} = \{877, 1755, 2633, 3510\}.$$

The prime $p < 10000$ with the largest $|\mathcal{E}_p|$ is $p = 2113$, and

$$\mathcal{E}_{2113} = \{44, 443, 553, 748, 1384, 1559, 1669, 2068, 2112\}.$$

We end with a table for the number of odd primes $p < 10000$ having the same number $|\mathcal{E}_p|$:

$ \mathcal{E}_p $	1	3	4	5	7	9	Total
$ \{3 \leq p < 2039\} $	192	92	1	18	4	0	307
$ \{2039 \leq p < 4547\} $	181	97	1	25	2	1	307
$ \{4547 \leq p < 7219\} $	187	101	0	17	2	0	307
$ \{7219 \leq p < 10000\} $	192	93	0	19	3	0	307
$ \{3 \leq p < 10000\} $	752	383	2	79	11	1	1228

Table 2

REFERENCES

- [1] D. W. Boyd, “A p -adic study of the partial sums of the harmonic series”, *Experimental Math.* **3** (1994), 287–302.
- [2] F. G. M. Eisenstein, “Eine neue Gattung zahlentheoretischer Funktionen, welche von zwei Elementen abhängen und durch gewisse lineare Funktional-Gleichungen definirt werden,” *Berichte Königl. Preuß. Akad. Wiss. Berlin* **15** (1850), 36–42.
- [3] A. Eswarathasan and E. Levine, “ p -integrals harmonic sums”, *Discrete Math.* **91** (1991), 249–257.
- [4] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 4-th edition, (Oxford), 1961.

353 FULWOOD ROAD, SHEFFIELD, S10 3BQ, UNITED KINGDOM

E-mail address: p.shiu@yahoo.co.uk